Saumya Sinha | Curriculum Vitae

1300 30th Street, Apartment B4 -16, Boulder, CO 80303

+7204211144 • saumya.sinha@colorado.edu
https://saumyasinha.github.io/

I completed my Masters and started as a CS PhD student at CU Boulder specializing in Machine Learning (ML). Prior to this, I have worked as a Data Scientist at American Express and Accenture for 3 years and have developed my expertise in predictive modeling and statistical analysis. At CU, I am a part of the Machine Learning Lab and I am interested in working on the intersection of computer vision and machine learning. Currently, I work on applying ML techniques for climate informatics.

Education

- **University of Colorado, Boulder** — **Boulder, CO**
  PhD in Computer Science, with Prof Claire Monteleoni — *2019 -*
- **University of Colorado, Boulder** — **Boulder, CO**
  Masters in Computer Science, GPA - 3.97/4 — *2017-2019*
- **Indian Institute of Technology, Kharagpur** — **Kharagpur, India**
  Integrated BS+MS in Mathematics and Computing, GPA - 8.4/10 (Department Rank - 5) — *2009–2014*

Publications

- **Detecting Avalanche Deposits using Variational Autoencoder on Sentinel-1 Satellite Imagery. (Spotlight)**
  Saumya Sinha, Sophie Giffard-Roisin, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni. *Tackling Climate Change with Machine Learning Neurips Workshop proceedings, Vancouver, Dec 2019*
- **Can Avalanche Deposits be Effectively Detected by Deep Learning on Sentinel-1 Satellite SAR Images?**
  Saumya Sinha*, Sophie Giffard-Roisin*, Fatima Karbou, Michael Deschatres, Anna Karas, Nicolas Eckert, Cécile Coléou, Claire Monteleoni.*Climate Informatics Workshop Proceedings, Paris, Oct 2019*

Professional service

- **Peer reviewer**: ICLR 2019 & 2020, Climate informatics 2019
- **Teaching**: Teaching Assistant for Data Strcutures (Fall 2019, Spring 2019, Fall 2018) , Graduate Course Assistant for Machine Learning (Spring 2018). Awarded the Outstanding TA award for the year 2018-2019.

Projects

CU Boulder

- **Learning to synthesize adaptive camouflage textures to hide an object by only making use of it's local background (in changing environments when one static camouflage is not enough.)**
  - Worked primarily on two major texture synthesizing models - data driven parametric approach (from the literature of style transfer) and a GAN-based method.
  - To measure the success, used a MaskRCNN detector to find out the extent to which we can adversarially attack them.
  - Used a simulation engine (Unreal) with high-fidelity 3D vehicle models to apply the camouflage pattern on the car's body and test it under a diverse set of environmental conditions.
- **Augmented Reality based IOS app using text detection and manipulation for people with reading disabilities.**
  - Worked on Augmented Reality development using ARkit.
  - Represented virtual 3D text obtained from an deep-learning based OCR model and anchored it over the real text.
  - Added hit test (single/double tap)functionality on the virtual 3D text to navigate to a new screen or display objects relevant to the text. Also enabled anchoring the objects to a plane and interacting with it through hand gestures.
- **Building a GAN based creative assistant for Logo Generation**

- Explored GAN generative models to create a creative assistance to help generate new and unique preliminary logo ideas.
- Generated a set of semantically meaningful clusters that could improve in conditional GAN training.
- Achieved best results (and mitigated mode collapse) with improved distance metric in the WGAN combined with ACGAN, conditioned on the above mentioned clusters.

- **Develop a scalable tool for Trend Classification in Twitter**
  - Built a Machine Learning pipeline with a Real-time Classifier which takes twitter data from a Kafka queue and writes into Cassandra. It was able to classify 6k tweets per minute with a pretty high accuracy(93%).
  - Tweets were classified into topics - Sports, Politics, Entertainment,Technology and Mood. Every trend was assigned a topic based on the majority classification of it's tweets.
- **Depth Map Estimation from Monocular Images**
  - Developed and compared various CNN based architectures from fully connected models to purely convolutional and also used transfer learning with pretrained models to estimate depth map from monocular images.
  - Used the depth map to identify and differentiate between the foreground and background of an image.
  - Visualised features formed by kernels.
  - Developed models on data with indoor scenes and objects. Achieved a very low RMSE on the object data, around 0.4.
- **Motion Planning with Rapidly Exploring Random Trees**
  - Explored and compared various variants of rapidly-exploring randomized-tree algorithm (RRT) to sub-sample the free space and find a path between two points in a 2D grid world with multiple obstacles.
  - Used GUI APIs (built by my team) to visualize and demonstrate the strengths and weaknesses of RRT, A-star and their variants like RRT*, Bi-directional RRT.

IIT Kharagpur

- Detected and Mined communities in twitter, implemented clustering and graph partitioning algorithms after exploring social networking graphs and various interactions between users.
- Implemented Collaborative Filtering Algorithm for recommending brands given a user's brand choice from a fairly large dataset of user-brand transactions.

Previous Employment

- **American Express** — **Bangalore, India**
  Data Analyst, Big Data Capabilities and Algorithms division — *March 2016–May 2017*

  Text Mining/Modelling to build merchant database
  - Worked to improve data quality of merchant (businesses using Amex services) attributes like merchant names and addresses in a consumer-friendly database by integrating data from miscellaneous information sources.
  - Implemented Bayesian Probabilistic model (baseline) and predictive models using GBM to choose most accurate value for a given attribute and achieved 95% accuracy for most of the markets.
  - Tested out a sequence to sequence learning model using recurrent neural networks (LSTMs) for guessing the correct or cleansed values for the above textual attributes.

  Developed Training Module
  - Created training modules on Python, R, Hive and Unix, trained a group of (20-30) people on the above languages/tools.

- **Accenture** — **Bangalore, India**
  Data Analyst — *June 2014 - Feb 2016*

  Predictive Maintenance
  - Worked for an Oil and Natural Gas resources client, creating reproducible code pipelines for failure detection of equipments in the plant, including data preprocessing, exploration, feature extraction, modelling, diagnosis, performance evaluation and visualization.
  - Wrote R code for Predictive modeling using Machine Learning techniques – GLM, SVM, Random Forest and Unsupervised Machine Learning using Markovian modelling (for cluster analysis).

Technical skills

- **Programming Languages:** Proficient in: Python, C++, big data tools – Hive, PySpark, Hadoop MapReduce
  Deep Learning Frameworks - Pytorch, Keras, Sklearn
- **Statistical Skills:** R, SQL, SAS